Monte Carlo Model of Reset Stochastics and Failure Rate Estimation of Read Disturb Mechanism in HfO_x RRAM

N. Raghavan[1, ♣], D.D. Frey[2], M. Bosman[3] and K.L. Pey[1]

[1]Engineering Product Development (EPD) Pillar, Singapore University of Technology and Design (SUTD), Singapore – 487 372.
[2] Department of Mechanical Engineering, Massachusetts Institute of Technology (MIT), Cambridge, MA 02139, USA.
[3] Institute of Materials Research and Engineering (IMRE), A*STAR, Singapore – 117 602.
♣Ph: (+65) 9862 1185, E-mail: nagarajan@sutd.edu.sg

Abstract — **Read disturb is a key failure mechanism in ultra-thin dielectric based HfO_x RRAM devices. While the physical origin of read disturb has been attributed to vacancy perturbations in the conductive filament due to high localized field around the filament constriction or at the dielectric tunnel barrier, there is no statistical model based on physical principles that enables quantification of the disturb failure probability (F_{DIST}) and the associated variability in the disturb voltage (V_{DIST}) for a given set of device operating conditions. This study develops a percolation based framework to model the filament configuration in the low resistance state (LRS) and a Kinetic Monte Carlo (KMC) routine to simulate the gradual evolution of the filament during the reset phase and identify the first voltage at which the disturb phenomenon is observed in the post-reset high resistance state (HRS). The model enables us to explain the bimodal nature of filament configuration after reset, wherein the filament may have ruptured (resulting in an oxide tunnel barrier) or it could have undergone shrinkage staying in the Quantum Point Contact (QPC) regime with a constriction that governs its stability. Two key external parameters, current compliance (I_{comp}) and voltage sweep ramp rate (RR) have been identified and their impact on the disturb failure distribution and reset kinetics has been examined in detail.**

Keywords – Conductive filament; Constriction; Quantum point contact; Read disturb; Tunnel barrier.

I. INTRODUCTION

Resistive random access memory (RRAM) technology has been the subject of intensive research over the past 6-8 years as it has several advantages over 3D NAND Flash technology which is reaching its limits in terms of scalability and reliability. Some of these advantages include simpler design and fabrication flow, CMOS-compatible material system, prolonged endurance and retention, aggressive scalability (demonstrated down to 10×10 nm^2 [**1**]), non-charge based robust data storage mechanism (with improved radiation resilience) and the potential to be integrated alongside logic devices for future hybrid system-on-chip (SoC) applications. Three key failure mechanisms in RRAM are endurance degradation, retention loss and read disturb [**2**]. While the basic understanding of these three mechanisms and their driving forces has been well established [**3**] – [**9**], there isn't enough effort devoted to developing quantitative failure probability models for these that could be used to estimate the device and memory circuit array lifetime at a desired percentile for a given set of operating conditions (SET and RESET voltage, pulse frequency, current compliance (I_{comp}), temperature etc…) and device dimensions (area and dielectric thickness) in order to assess the commercial viability and fabrication quality of the technology in the foundry. Though some initial efforts to model endurance and retention failure statistics have been made [**10**, **11**], read disturb mechanism has yet to be modeled from a physical-statistical viewpoint. While a detailed statistical study on the disturb phenomenon has been carried out recently by Luo et al. [**12**], it does not correlate the failure probability to physical parameters of the device, which limits its practical value for reliability design. Since the spread in disturb failures also depends on the spread in the reset state configuration of the filament, a holistic read disturb study has to begin from looking at the reset sweep kinetics.

In this study, we begin from the low resistance state (LRS) and simulate the gradual time and voltage governed shrinkage of the filament during the reset sweep. At the end of the reset sweep, the shape and size of the filament is noted and these two factors subsequently govern the next stage of simulation to estimate the voltage at which the filament would be perturbed (V_{DIST}) in the high resistance state (HRS). The focus here is on oxygen vacancy (V_O^{2+}) - (O^{2-}) oxygen ion based metal-insulator-metal (M-I-M) HfO_x RRAM, where the filament comprises of a collection of oxygen vacancies and the LRS-HRS transitions are determined by the drift of oxygen ions which can get stored in the top/bottom electrodes depending on the polarity of the applied voltage. The SET stage involves field induced generation of vacancies due to Hf-O bond breaking, while the RESET stage involves drift of oxygen ions from the electrode into the oxide and recombination with the relatively immobile vacancies. Our analysis shows that the local potential / field distribution in the dielectric in HRS with partially ruptured / shrunk filament and the variance in its shape / size has a large influence on mean and spread of V_{DIST}.

The structure of this work is organized as follows. In Section II, the overall simulation framework and theoretical construct that has been developed to model the RESET and disturb phenomenon is explained. Section III shows the results of the RESET simulation in terms of the HRS resistance (R_{HRS})

distribution, evolution of filament resistance (R_{FIL}) during the reset voltage sweep as well as the resulting bimodal nature of the filament configuration. The statistical trends of V_{DIST} and their dependence on I_{comp} and the voltage ramp rate (RR) are presented in Section IV. From a material design perspective, Section V briefs on the influence of the activation energy for vacancy generation (E_{aG}) as well as the field acceleration factor (γ) (which are unique to every high-κ material) on the distribution of V_{DIST} and finally, Section VI summarizes the outcomes and inherent assumptions of this study and concludes with recommendations for further research along this line.

II. Proposed Simulation Framework

Our simulation begins from the LRS state, focusing on the reset voltage sweep and tracking the gradual evolution of the conductive filament (CF) during this sweep. Fig. 1 is a flow chart that describes the procedure followed. The complete simulation flow shall be explained in different stages in the following sub-sections.

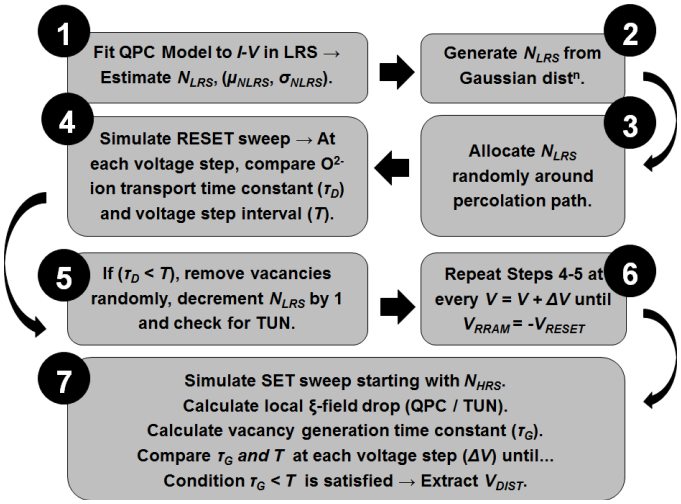

Figure 1. Flowchart describing the step-by-step process of simulating the kinetics of the reset process (LRS → HRS transition) and the read disturb phenomenon in the HRS in order to extract the statistical trends of evolving filament resistance (R_{FIL}) and disturb voltage (V_{DIST}).

A. Initial Vacancy Count in the LRS

Since our framework is based on the percolation theory, the number of initial vacancies / defects (N_{LRS}) in the filament should be known prior to carrying out the RESET simulation. When considering multiple switching cycles, this value of N_{LRS} will have a Gaussian distribution with a mean and standard deviation of μ_{NLRS} and σ_{NLRS} respectively (as documented in Ref. [**13**]). For simplicity, we assume the values of {μ_{NLRS}, σ_{NLRS}} here as a function of different I_{comp}. Logically, the value of μ_{NLRS} should increase and σ_{NLRS} decrease for higher I_{comp} (equivalent to harder breakdown of the dielectric which results in more vacancy creation and less variation in their count). The values chosen for our simulation are tabulated in Table I. A more rigorous approach to address this would be to model the measured I-V curve for each device in the LRS with the Quantum Point Contact (QPC) model, as proposed by Degraeve *et al.* [**14**, **15**], which can always be used for any RRAM device in the SET state. With different I-V curves corresponding to different values of N_{LRS} based on the QPC formulation, using the current conduction level and the fluctuations (ΔI) in the current in the I-V data, an approximate value for N_{LRS} could be deduced assuming a certain shape for the filament as determined by the quantum energy band frequency parameters. For more details on the estimation of defect count, readers are advised to access Ref. [**13**].

TABLE I. Values of Gaussian distribution parameters chosen for defect count in LRS state for the Monte Carlo simulation

Compliance Current (I_{comp})	μ_{NLRS} *(#)*	σ_{NLRS} *(#)*	σ_{NLRS} / μ_{NLRS} *(No Units)*
100μA	80	5	6.25%
50μA	50	8	16.0%
25μA	25	6	24.0%
10μA	15	7	46.7%
1μA	11	6	54.5%

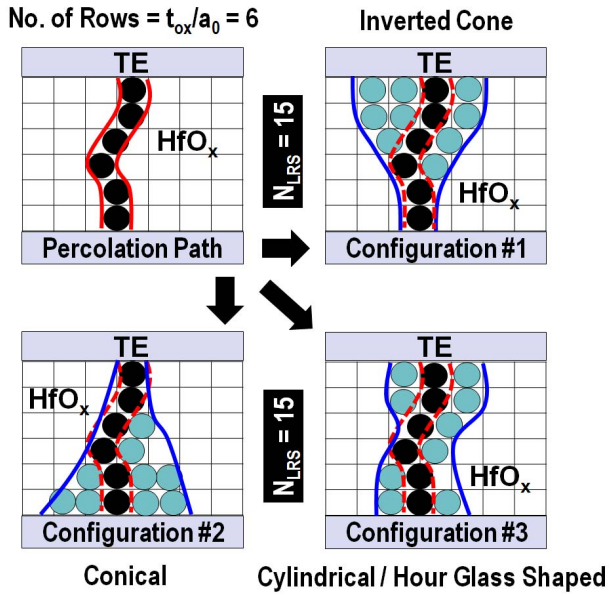

Figure 2. Illustration of the percolation cell matrix showing the placement of defects (black dots) to first create a percolation path (conductive filament) in the dielectric followed by three equally probable vacancy cluster configurations (N_{LRS} = 15 defects) representing filaments that are inverted cone, normal cone and cylindrical / hour-glass in shape. The Monte Carlo code developed in this work does not provide any means to assess the thermodynamically preferred shape and vacancy arrangement of the filament. The blue dots represent the additional defects placed around percolation path.

B. Configuring the Filament in the LRS

With the number of defects known in the LRS (extracting a value of N_{LRS} from its Gaussian distribution), their arrangement also plays a key role. We first assign one defect each to every single row of the 2-D percolation cell matrix (see Fig. 2(a)) along a certain vertical connected column. For an oxide with thickness t_{ox} and defect size a_0 (~6Å for HfO_2), this corresponds to t_{ox} / a_0 defects. The remaining defects (N_{LRS} - t_{ox}/a_0) are then allocated without any vertical spatial preference all along the CF next to the percolation column (refer to Figs. 2(b)-(d) for illustration with N_{LRS} = 15) which means that the

final shape of the filament could be conical, inverted cone, cylindrical or hour-glass in shape. While it is necessary to determine the right shape of the filament, there is no clear understanding on the driving forces that cause the filament to be asymmetric; as a result, we refrain from speculating the preferred shape and explore switching variability more conservatively by giving equal weightage to the nucleation of the three filament shapes shown in Fig. 2.

C. Simulation Procedure for Reset Sweep

The two main processes during the reset are defect transport and defect recombination. In general, recombination is a fast event, with a very high rate of occurrence [**16**]. As a result, the only rate-determining step for reset is the transport of O^{2-} ions, which can be described by the Mott's ionic transport model [**17**]. The rate of O^{2-} drift can be expressed by Eqn. (1), where τ_D is the time constant for drift, λ_D is the drift rate, ν is the vibration frequency (~10^{13} s^{-1}), d is the average drift distance into the dielectric, a is the interatomic distance (~1 nm) [**18**], ξ is the electric field, E_{aD} (assumed = 0.8eV [**19**]) is the activation energy for ion transport, q is the electronic charge, k_B is the Boltzmann constant and T_{FIL} is the internal filament temperature, which is in turn expressed by Eqn. (2) at steady state. The symbols T_0, V, R_{TH} and R_{FIL} refer to the ambient temperature, applied voltage, thermal resistance of the dielectric and the effective resistance of the conductive filament. The second term to the right of Eqn. (2) is in effect due to Joule Heating. The value of R_{FIL} is computed by considering the defects in a single row to be in "parallel" configuration and the defects in different rows to be in a "series" pattern. The arrangement and location of the defects and their connectivity determines the computed value of R_{FIL} at each stage of the simulation.

$$\frac{1}{\tau_D} = \lambda_D = \frac{\nu}{d} \cdot a \cdot \exp\left(-\frac{E_{aD}}{k_B T_{FIL}}\right) \cdot \sinh\left(\frac{qa\xi}{2k_B T_{FIL}}\right) \quad (1)$$

$$T_{FIL} = T_0 + \frac{R_{TH}}{R_{FIL}} \cdot V^2 \quad (2)$$

The reset sweep is simulated based on the procedure outlined in Fig. 1. At each voltage step of the *I-V* sweep, each oxygen ion is associated with a random value of time, *t'*, extracted from the exponential distribution with time constant (τ_D) and compared with the time duration at that voltage (Δt). If $t' < \Delta t$, the ion recombines with a vacancy and one vacancy is annihilated (removed) from the filament, without any spatial preference (vacancy passivation assumed to be equally probable at all locations of the filament). This same procedure is recursively carried out for each remaining oxygen ion at every voltage step starting from $V = 0V$ to $-V_{RESET}$. At the end of the voltage sweep, the filament is examined for any tunnel barrier and classified as TUN or QPC (connected filament) accordingly. Note that during this reset sweep, the temperature dynamically changes as the filament shape and size change and applied voltage increases as well.

D. Simulation Procedure for Disturb Voltage Estimation

After reaching the reset state, the disturb event during memory read operation is triggered by the possible generation of an additional defect in the ruptured / shrunk filament due to electric field and/or temperature. The voltage needed to SET (V_{SET}) is typically quite high compared to V_{DIST} [**12**] and it is the local electric field in the reset state that tends to perturb the filament quite a bit. Although it is difficult to derive an accurate expression for the electric field in the dielectric at the ruptured filament region (as the filament shape and size are not easy to decipher), we may formulate some simple logical / empirical expressions for the same as a first order approximation. With this in mind, the maximum field in the TUN and QPC modes may be expressed as ξ_{TUN} and ξ_{QPC} in Eqns. (3) and (4) respectively where t_{BAR} is the tunnel barrier thickness and $\{\alpha, \Delta\}$ are arbitrary constants that determine the potential gradient at the filament-constriction interfaces in the case of a QPC configuration. It is qualitatively established that the sharper (more abrupt) the interface between the filament and constriction, the higher the local field would be, as documented in Ref. [**20**], wherein the superposition of the incident and reflected electron wavefunctions is believed to be the key to the creation of an enhanced field locally at the interfaces [**21**]. While the average ξ-field ($\xi_0 = V_{READ}/t_{ox}$) may be quite low in these devices at read voltage conditions of 100mV, the partial rupture / shrinkage of the filament causes the field to be locally intense that can cause unintended bond breaking and V_0 generation. Note in Eqn. (4) that for a large defect count, the second term becomes negligible meaning that the local field tends towards the global value of ξ_0 resulting in Ohmic-like behavior for LRS with higher I_{comp}. Therefore, the expression in Eqn. (4) equation accounts for both QPC and Ohmic operation regimes (it is however valid only for the case of $N_c \geq t_{ox}/a_0$, when a connected network of defects exist).

$$\xi_{TUN} = \frac{V}{t_{BAR}} \quad (3)$$

$$\xi_{QPC} = \xi_0 + (\alpha - 1) \cdot \xi_0 \cdot \exp\left(-\frac{N_C}{\Delta}\right);\ \xi_0 = \frac{V}{t_{ox}} \quad (4)$$

The method for simulating the disturb event is analogous to the reset case. We begin from V_{TE} = 0V and increment the voltage by ΔV after every time interval Δt, depending on the ramp rate (TE: top electrode). The expression for the rate of defect generation, λ_G is given by Eqn. (5) based on the thermochemical model [**22**, **23**], where E_{aG} (assumed = 1.1eV) is the activation energy for Hf-O bond breaking, γ is the field acceleration factor, p_0 is the dipole moment and κ is the relative permittivity of the dielectric. With a randomly chosen value of generation time (*t'*) from the exponential distribution with time constant (τ_G), the very first voltage level at which $t' < \Delta t$ is classified as V_{DIST}.

$$\frac{1}{\tau_G} = \lambda_G = \nu \cdot \exp\left(-\frac{E_{aG} - \gamma \cdot \xi}{k_B T_{FIL}}\right);\ \gamma = p_0 \cdot \left(\frac{2+\kappa}{3}\right) \quad (5)$$

The procedure outlined above is performed multiple times (100-500 simulation trials) in order to extract the distribution of various parameters such as V_{DIST} and R_{HRS} at multiple I_{comp} levels. As a standard stack, we consider an oxide with six rows of defects ($t_{ox} = 6a_0$ = 3.6 nm).The resulting statistical trends will now be presented in the following sections.

III. Filament Evolution During Reset

The dynamic evolution of R_{FIL} for multiple reset cycles is shown in Fig. 3 for the case of I_{comp} = 50µA and RR = 100mV/sec. In computing the value of R_{FIL}, the effective resistance of the presence and absence of a defect in a cell belonging to the percolation matrix is assumed to be R_{defect} = $10^5\Omega$ and $R_{no\text{-}defect}$= $10^7\Omega$ respectively. Note that the resistance jumps are initially small while in the QPC regime and they gradually increase in magnitude as the filament shrinks. For high V_{RESET} close to 1.25 – 2V, abrupt jumps in R_{FIL} by more than an order of magnitude are observed (as shown in the inset). This logarithmic change corresponds to a transition from the QPC to the TUN mode of conduction as the CF ruptures and a tunnel oxide barrier is introduced at some location along the filamentary path. The corresponding changes in T_{FIL} during these reset sweeps are captured by our simulation and plotted in Fig. 4 (based on the expression in Eqn. (2) where R_{TH} for HfO_2 is assumed to be 2 × 10^5K/W). While the value of T_{FIL} peaks and saturates if the filament remains in the QPC mode throughout the sweep (accumulative Joule Heating and current confinement during filament "necking"), the temperature jumps down considerably towards 300K (T_0) for TUN mode. The insets in the figure clearly show this distinction in the temperature profiles.

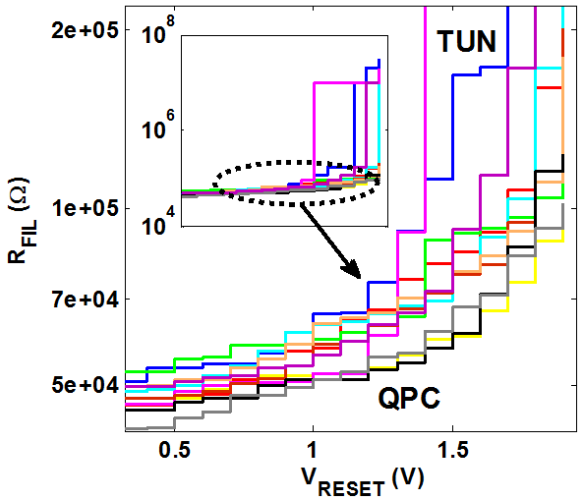

Figure 3. Evolution of R_{FIL} during the reset sweep for multiple simulation trials considering I_{comp} = 50µA and RR = 100mV/sec. The logarithmic jumps in R_{FIL} correspond to the rupture of the filament and a transition from the QPC to TUN mode of conduction.

The distribution of R_{FIL} when the reset sweep is arrested at different stages is presented in Fig. 5 on a lognormal scale. Note the increasing spread in the distribution (and deviation from lognormal trend as also evident from variability characterization results on HfO_2 films in Ref. [**24**]) as the maximum reset stop voltage ($V_{RESET\text{-}STOP}$) is increased from -0.50V to -1.50V. The trend of R_{FIL} is not perfectly lognormal possibly because the resulting values of R_{FIL} are a convolution of the stochastics of the filament size and shape in LRS as well as the stochastics of the defect annihilation process during the reset sweep. The combined effect of these factors results in a non-standard distribution for R_{FIL}. While the main plot in Fig. 5 only shows the case of QPC filaments, the inset also includes the TUN mode resistance data that becomes more probable with higher $V_{RESET\text{-}STOP}$.

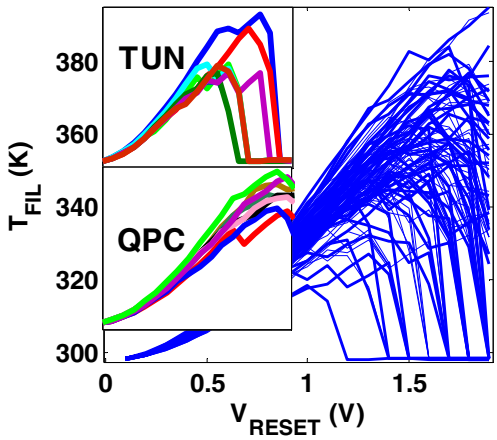

Figure 4. Corresponding profile of filament temperature (T_{FIL}) change during the reset sweep for the filament evolution represented in Fig. 3.

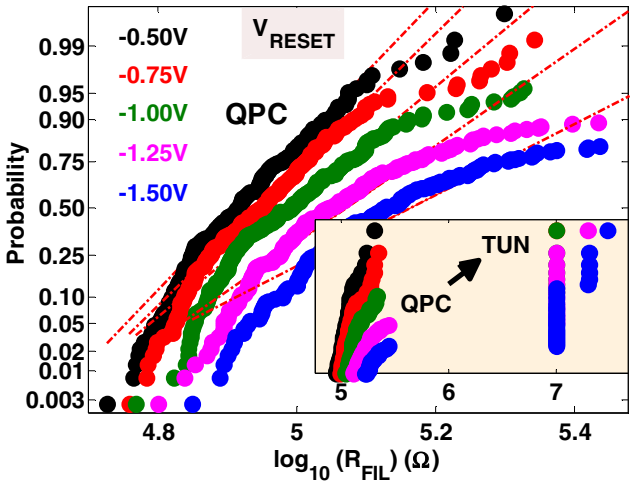

Figure 5. Lognormal distribution plot of R_{FIL} for different reset stop voltage values ranging from -0.50V to -1.50V. The inset shows the TUN event data as well, while the main figure only plots the distribution for the QPC configuration of the filament in the reset state. The deviation from the standard lognormal behavior is also evident in the electrical characterization results presented in Ref. [**24**].

It is obvious from the results presented that the choice of $V_{RESET\text{-}STOP}$ is a very critical one that impacts RRAM switching reliability and variability. While a very low or very high $V_{RESET\text{-}STOP}$ would help in reducing the spread of R_{FIL} by ensuring most of the filaments either end up in QPC or TUN mode, the memory window could be substantially low. Too high a value of $V_{RESET\text{-}STOP}$ could end up creating more additional vacancy defects than what it can passivate, affecting the dynamic balance between defect generation and recombination, thereby causing memory window to shrink and also increasing the risk of a permanent hard breakdown of the stack. The apparent bimodality in filament configuration seems to be a detrimental factor in making resistance variability control a difficult option for RRAM devices. Fig. 6 schematically explains this issue (using the results of two different Monte Carlo simulation trials) whereby for the same N_{LRS} = 14 defects to begin with in the LRS state with different filament shapes, passivation of

only three vacancy defects at different locations can give rise to totally different HRS configurations → one in QPC and the other in TUN mode.

To further examine what percentages of devices / switching cycles end up in the QPC and TUN modes, a detailed investigation involving 25 repeated trials of 100 switching cycles (reset transition) each was carried out at five different values of I_{comp} = {1, 10, 25, 50, 100}µA. The proportion of QPC and TUN events (p_{QPC}, p_{TUN}) was extracted for each trial and the mean and spread for the 25 trials was computed and plotted in Fig. 7 as a stem plot. It is interesting to note that even for a high I_{comp} = 100µA, there is a finite chance of 3-8% that the filament could end up in TUN mode. As for I_{comp} = 1µA, there is equal chance that the resulting filament could either be in QPC or TUN.

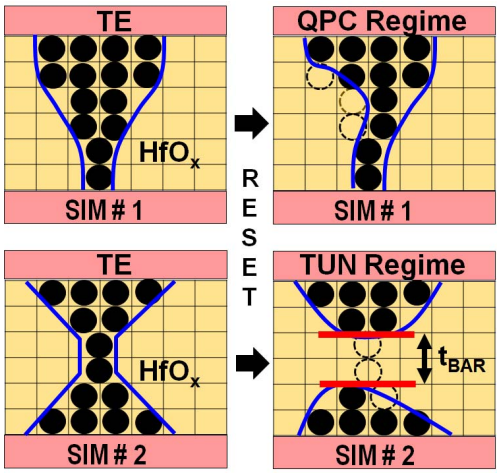

Figure 6. Illustration from the Monte Carlo simulation where two filaments in LRS with same N_{LRS} value (but different shape) end up in QPC and TUN modes after the removal of three defects each during the reset sweep. The dotted circles represent the passivated vacancies.

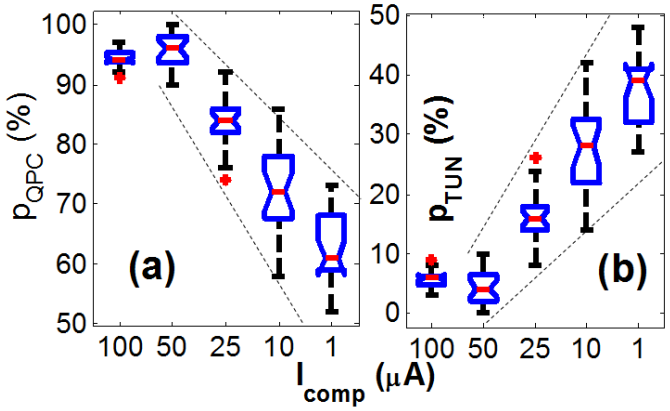

Figure 7. Trend of percentage of devices / switching cycles that end up in the (a) QPC and (b) TUN modes respectively as a function of I_{comp}. This stem plot shows the median and the 25th, 75th percentile values for p_{QPC} and p_{TUN}.

Since we use the percolation network to visualize the filament, the thickness of the tunnel barrier and the width of the QPC constriction can both be digitally extracted from the simulation results as a function of the defect size, a_0. In Fig. 8, we show a bar graph that plots the percentage of reset cycles that end up with QPC and the percentage that end up in a TUN mode with integer barrier thickness of $1 \times a_0$, $2 \times a_0$ and $3 \times a_0$ for three different I_{comp} values = {1, 10, 25}µA. At low I_{comp}, there is a finitely high chance of having a TUN barrier that is 2-3 cells thick, which implies that the reset could be quite deep and subsequent value of V_{DIST} very high.

In Fig. 9, the lognormal plot of final filament resistance in the reset state is shown for different I_{comp} (only QPC configuration is plotted). Again, just as the spread in distribution increases for higher $V_{RESET\text{-}STOP}$, the spread is higher for lower I_{comp} (which is logically expected). While ultra-low switching power is desirable in RRAM from a performance point of view for which we have to operate it at low I_{comp}, this becomes a problem in terms of variability. Therefore, reliability and performability do not go hand in hand and an acceptable trade-off between these two factors should be considered depending on the material technology and its intended application. In the next section, we shall look into the statistical trends of V_{DIST} as extracted from our modeling framework.

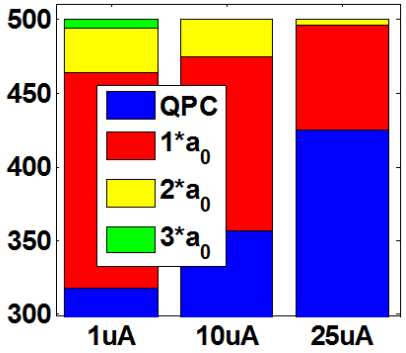

Figure 8. Bar graph showing the number of simulation trials (totaling 500) that end up in QPC (blue), $1 \times a_0$ tunnel barrier (red), $2 \times a_0$ tunnel barrier (yellow) and $3 \times a_0$ tunnel barrier (green) for three different I_{comp} ~ {1, 10, 25}µA. Note that the y-axis only begins at 300 as all three I_{comp} values have atleast 300 QPC based filament configurations.

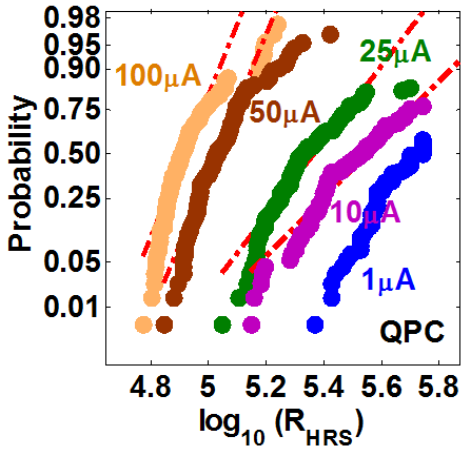

Figure 9. Final filament resistance distribution (lognormal scale) after reset sweep for a wide range of I_{comp}, keeping $V_{RESET\text{-}STOP}$ fixed at -2V. Only the QPC configured filament resistances are shown in this plot.

IV. Estimation of Disturb Voltage Statistics

Using the procedure outlined in Section II.D, the value of V_{DIST} is estimated for 100 simulation trials each, at I_{comp} ranging from 1-100µA, considering RR = 10mV/sec. The distribution

plot for V_{DIST} considering only the QPC scenario is shown on a Weibull scale in Fig. 10. With higher I_{comp} (↑ N_{LRS} ► ↑ N_{HRS}), the mean value of V_{DIST} increases significantly as the electric field at the filament-constriction interface drops considerably. The higher variance in N_{HRS} at low I_{comp} translates to greater variance in V_{DIST} trends as well. With the standard read voltage for memory devices being V_{READ} = 100mV, we may deduce from Fig. 10 that at low percentiles for I_{comp} ~ 1-10μA, there is a finitely high possibility of the filament being perturbed during the low voltage read process by generation of additional V_0 defects. However, if the I_{comp} is chosen to be high enough, then the risk of disturb events is considerably reduced. Fig. 11 illustrates the field enhancement that could occur in a QPC configuration as a function of the number of defects in the filament at HRS, for assumed values of α = 10 and Δ = 5 (according to Eqn. (4)). When compared to the standard ξ-field for the dielectric (which is valid for $N > 25$), the enhancement due to QPC effect can be as much as 2-3 times of the standard value for N ranging from 6 to 15.

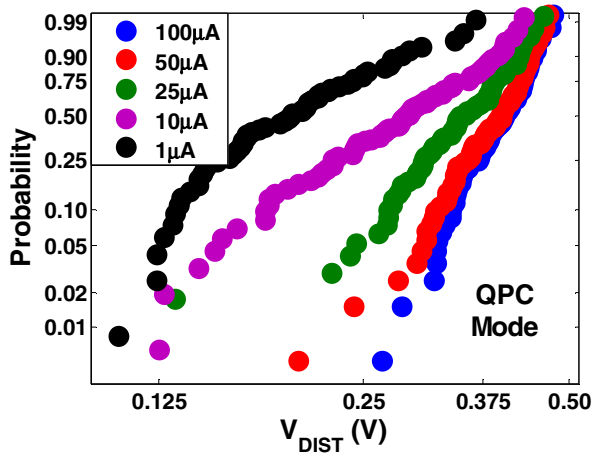

Figure 10. Distribution plot of V_{DIST} at different I_{comp} values for filaments that remained in the QPC mode after reset sweep.

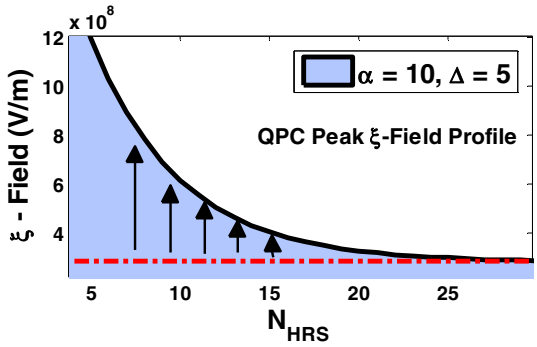

Figure 11. Emprical (logical) trend of localized ξ-field in the QPC mode configuration (at the filament-constriction interface) as a function of the number of defects in the filament at HRS (N_{HRS}).

The trend of mean value of V_{DIST} and its Weibull shape parameter (in QPC mode) for different I_{comp} values is plotted in Fig. 12(a) and (b) respectively. As already evident from Fig. 10, the V_{DIST} distribution gets sharper (↑ Weibull slope (β)) and larger in value (↑η) with increase in I_{comp} setting. As for the TUN mode, the mean value shows the opposite trend (Fig. 12(c)) because the average thickness of the tunnel barrier for wider filaments is likely to be much smaller. As a result, the applied voltage drops across a very thin barrier, resulting in high applied field and hence lower V_{DIST} at higher I_{comp}.

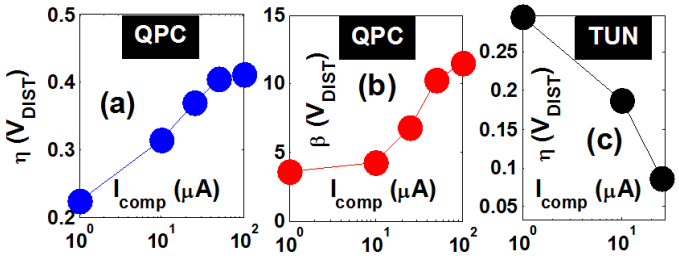

Figure 12. (a, b) Mean and spread (scale and shape parameter) of the V_{DIST} Weibull distribution as a function of I_{comp} for filaments in the QPC mode (RR = 100mV/sec). (c) Trend of mean V_{DIST} with I_{comp} for filaments in the TUN mode.

The dependence of the scale and shape parameters for V_{DIST} as a function of the RR is shown in Fig. 13, again for the QPC mode, keeping I_{comp} fixed at 100μA. We see a clear increase in V_{DIST} as well as a reduction in the variance (↑β) for a higher RR. This implies that operating the RRAM device at high frequency pulsing conditions can enhance disturb immunity while reading when compared to the case of DC / low frequency. At higher ramp rates, there is relatively less time (lower probability) at each voltage step to induce vacancy perturbations and as a result, the value of V_{DIST} tends to naturally increase with RR. For very high RR, the value of V_{DIST} could well approach the value of V_{SET}.

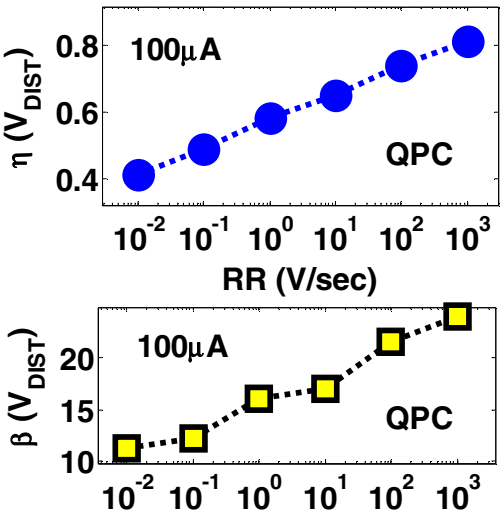

Figure 13. Mean and spread (scale and shape parameter) of the V_{DIST} Weibull distribution in the QPC configuration mode as a function of ramp rate (RR) for a fixed I_{comp} = 100μA.

Since both I_{comp} and RR have an impact on the probability of read disturb, a 3D plot is presented in Fig. 14, where the disturb probability ($F_{DIST} = Pr(V_{DIST} < V_{READ})$) is computed for a wide range of both I_{comp} and RR values, given the information on the Weibull parameters for V_{DIST} in the QPC and TUN modes and with the knowledge of p_{QPC} and p_{TUN}. The expression for F_{DIST} can be given by Eqn. (6) as a linear superposition of the failure probabilities in the QPC and TUN

configurations, weighted by their bimodal proportion weights, which we have already estimated previously in Fig. 7.

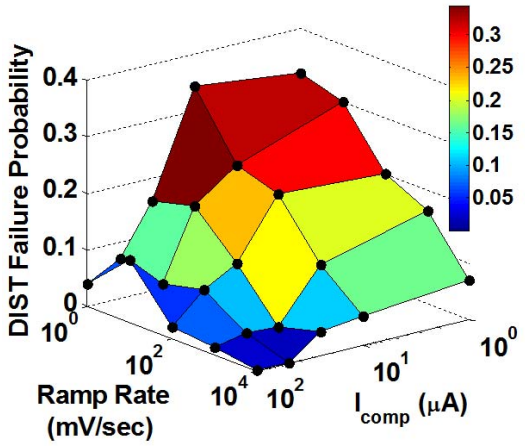

Figure 14. 3D plot of the disturb failure probability (F_{DIST}) dependence on *RR* and I_{comp}, accounting for the bimodal configuration of the filament in the HRS.

Given that *RR* and I_{comp} are both user-defined parameters for RRAM operation, the graph in Fig. 14 can be a useful tool to have a quantitative estimate of the failure probability due to read disturb events. The model presented here can hence be used as a design for reliability tool in the foundry for RRAM stack qualification and for optimizing the set of operation parameters to address the performance – reliability trade-off.

$$F_{DIST} = \frac{p_{QPC}}{p_{QPC}+p_{TUN}} \cdot \Pr\left(V_{DIST}^{\ QPC} < V_{READ}\right) + \frac{p_{TUN}}{p_{QPC}+p_{TUN}} \cdot \Pr\left(V_{DIST}^{\ TUN} < V_{READ}\right) \quad (6)$$

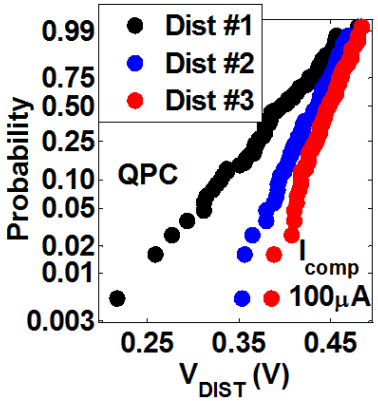

Figure 15. Shift in the V_{DIST} distribution for the case of multiple successive disturb phenomena. As the number of disturb events the device can withstand increase, the distribution gets tighter. The simulation here was carried out with I_{comp} set to 100μA.

It is likely that in many cases, the very first disturb event may not be the single detrimental one that affects the memory state (filament in QPC mode). Rather, the disturb phenomenon could be a cumulative effect of multiple vacancy perturbations in the CF. We illustrate one such scenario in Fig. 15 which plots the distribution of the first three consecutive disturb events in a device that is gradually stressed in the HRS. It is worth noting that there is a significant rightward shift in the distribution of V_{DIST} at low percentiles and the distribution gets tighter for more disturb events. This is because the local QPC-field reduces substantially when the defect count (N_{HRS}) is incremented by one every time a disturb event occurs. After a few disturb events, the value of ξ-field approaches the classical value of V/t_{ox} and therefore, the 'almost' uniform field in the multiple simulation trials causes the V_{DIST} distribution to tighten for subsequent disturb.

V. Material Design for Disturb Immunity

From a materials perspective, the read disturb events are heavily influenced by two key parameters of the dielectric which govern the value of $\lambda_G \rightarrow$ they are E_{aG} and γ, where γ is directly proportional to p_0 and κ for the high-κ film. By choosing different dielectric materials (or dopants) other than just undoped HfO_2, we may easily tune the values of E_{aG} and γ to optimize the reliability of the RRAM device. In Fig. 16, the impact of changing only one of these two parameters and changing both of them at once on disturb immunity (V_{DIST}) is illustrated by choosing arbitrary values for $\{E_{aG}, \gamma\}$ in the simulation code. It is observed that increasing γ alone results in a shift of V_{DIST} from the blue to red dot pattern with a significant drop in immunity, though the spread in the distribution is now well reduced. Since γ represents the sensitivity of the dielectric chemical bond strength to the local ξ-field, increase in γ alone does not serve our need. When the E_{aG} value is also simultaneously increased, the distribution shifts from the red to the black dot pattern with approximately the same variance. Now, the device has very good immunity as well as low spread in disturb voltage trends. Hence, increasing $\{E_{aG}, \gamma\}$ together helps in making the RRAM more robust to disturb events. This requires a careful choice of the right dielectric materials and dopants.

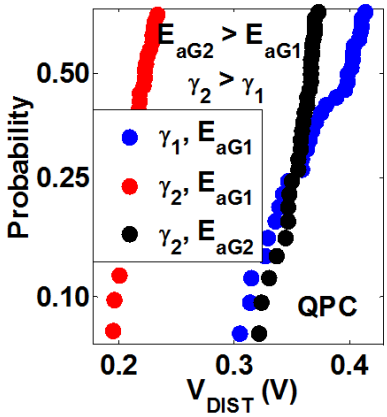

Figure 16. Shift in the V_{DIST} distribution (QPC mode filament) when the bond breaking activation energy (E_{aG}) and/or field acceleration factor (γ) for the dielectric are altered.

VI. Conclusion

A detailed study on the stochastics of the reset phenomenon and read disturb failures has been presented in this work using a Monte Carlo based methodology and a percolation based geometric framework. The study helped in quantifying the impact of various material and operational factors (I_{comp},

$V_{RESET-STOP}$, RR, E_{aG}, γ) on the distributions of filament resistance (R_{FIL}, R_{HRS}) and disturb voltage (V_{DIST}). We were able to prove using our physical modeling approach that the filament configuration is almost always bimodal and the proportion of filaments ending up in the QPC or TUN mode is a strong function of I_{comp}. The detailed method of analysis prescribed here can be implemented in the foundry as a useful tool for robust design and engineering of the metal-insulator-metal (M-I-M) stack. Our approach also takes into consideration the role of quantum confinement and the unique localized potential gradient profiles that are specific to the QPC configuration. The presence of a non-uniform field (more ξ-field concentrated at the filament-costriction interface) complicates the statistical process quite a bit. During the SET stage, the field in the filament would dynamically evolve as a function of the V_0 defect count. This is unlike other models where the field is considered constant during the switching process. Although the approach we have taken here is holistic and generic and easily applicable to any oxygen-based RRAM (OxRAM) device technology, there are many simplifying assumptions we have made along the way that deserve more investigation in the future versions of the model.

First of all, we placed defects randomly in the filament for LRS without any preference along the vertical axis. This makes it equally probable to have a conical and inverted cone filament. From experimental studies though, the shape of the filament is either one but not both [**25**, **26**]. The influence of the top and bottom electrodes, their symmetry/asymmetry in material, physical thickness and oxygen gettering capacity will all have an effect on shaping the CF in the LRS. The impact of nearby vacancies could also influence the generation of a new vacancy in their proximity and this could also dictate the shape of the filament. All these factors have to be dealt with analytically.

The second constraint in our study is the assumption of an average drift distance for all the O^{2-} ions residing in the top electrode (TE) to reach the vacancies for recombination. Since different vacancies in the filament are located at different distances from the TE-dielectric interface, the possibility of ion-vacancy recombination during the reset should specifically account for this distance, instead of assuming it to be an average value. Thirdly, the formulation for the electric field in the QPC mode is empirical in our study. Although it is hard to derive an analytically straightforward expression for the spatial variation of the potential along a QPC, the model here has to be modified to make it more physical. The location of the constriction in the CF and the width of the filament-constriction interface also have to be accounted for. Studies on these factors are still lacking in the RRAM community and further in-depth theoretical analysis would be needed.

Next, the temperature in the filament was assumed to be uniform all around, which is a first-order approximation. The exact profile of the temperature [**27**] along the x- and y-axis of the CF has to be derived and incorporated into the model. Whether the QPC scenario and quantum confinement effects cause changes to the temperature at the constrictions is also not clear at the moment. The local changes in temperature can be a key factor affecting the voltage and resistance distribution and the classical viewpoint we have taken cannot address these intricacies. Also, to keep things simple, we only considered a 2D scenario of percolation here, while the CF is in reality a 3D object. Extending the analysis here to the third dimension (keeping the percolation framework) will certainly help in making the model one step closer to reality. The role of the dielectric-electrode interfaces in serving as a diffusion barrier, the possible lateral diffusion (and depletion) of oxygen ions / vacancies and the enhancement in electric field (if any) at the tips of ruptured filaments (in TUN mode) were also not considered here. With many unanswered questions and simplifications listed above, there is still a lot of scope for further improvement of the model.

Since the principle of switching in conducting bridge RAM (CBRAM) (metal filamentation by redox chemistry) is very different from the case of OxRAM we have considered here, a completely different framework would be needed to analyze the CF formation and evolution in CBRAM devices [**28**]. This is one focal point of our future study. Extension of the Monte Carlo model we have developed to look into retention failure statistics of OxRAM at both resistance states is currently in progress.

ACKNOWLEDGMENT

The authors would like to thank the International Design Center (IDC) at the Singapore University of Technology and Design (SUTD) for fully funding this research work and provision of the travel allowance under Grant No. IDG11300103.

REFERENCES

[1] B. Govoreanu, G.S. Kar, Y.Y. Chen, V. Paraschiv, S. Kubicek, A. Fantini, I.P. Radu, L. Goux, S. Clima, R. Degraeve, N. Jossart, O. Richard, T. Vandeweyer, K. Seo, P. Hendrickx, G. Pourtois, H. Bender, L. Altimime, D.J. Wouters, J.A. Kittl and M. Jurczak, "10×10 nm^2 Hf/HfO_x crossbar resistive RAM with excellent performance, reliability and low-energy operation", *IEEE International Electron Devices Meeting (IEDM),* pp.31.6.1 - 31.6.4, (2011).

[2] N. Raghavan, "Performance and reliability trade-offs for high-κ RRAM", *Microelectronics Reliability*, Vol. 54, Issue 9, pp.2253-2257, (2014).

[3] Y.Y. Chen, R. Degraeve, S. Clima, B. Govoreanu, L. Goux, A. Fantini, G.S. Kar, G. Pourtois, G. Groeseneken, D.J. Wouters and M. Jurczak, "Understanding of the endurance failure in scaled HfO_2-based 1T1R RRAM through vacancy mobility degradation", *IEEE International Electron Devices Meeting (IEDM),* pp.20.3.1-20.3.4, (2012).

[4] B. Chen, J.F. Kang, B. Gao, Y.X. Deng, L.F. Liu, X.Y. Liu, Z. Fang, H.Y. Yu, X.P. Wang, G.Q. Lo and D.L. Kwong, "Endurance degradation in metal oxide-based resistive memory induced by oxygen ion loss effect," *IEEE Electron Device Letters*, Vol. 34, No. 10, pp. 1292-1294, (2013).

[5] B. Chen, Y. Lu, B. Gao, Y.H. Fu, F.F. Zhang, P. Huang, Y.S. Chen, L.F. Liu, X.Y. Liu, J.F. Kang, Y.Y. Wang, Z. Fang, H.Y. Yu, X. Li, X.P. Wang, N. Singh, G.Q. Lo and D.L. Kwong, "Physical mechanisms of endurance degradation in TMO-RRAM", *IEEE International Electron Devices Meeting (IEDM)*, pp.283-286, (2011).

[6] Y.Y. Chen, L. Goux, S. Clima, B. Govoreanu, R. Degraeve, G.S. Kar, A. Fantini, G. Groeseneken, D.J. Wouters and M. Jurczak,

"Endurance/Retention Trade-off on HfO_2 / metal cap 1T1R Bipolar RRAM", *IEEE Transactions on Electron Devices,* Vol. 60, No. 3, pp.1114-1121, (2013).

[7] N. Raghavan, R. Degraeve, A. Fantini, L. Goux, S. Strangio, B. Govoreanu, D.J. Wouters, G. Groeseneken and M. Jurczak, "Microscopic origin of random telegraph noise fluctuations in aggressively scaled RRAM and its impact on read disturb variability", *IEEE International Reliability Physics Symposium (IRPS)*, pp.5E.3.1-5E.3.7, (2013).

[8] B. Gao, J.F. Kang, H.W. Zhang, B. Sun, B. Chen, L.F. Liu, X.Y. Liu, R.Q. Han, Y.Y. Wang, Z. Fang, H.Y. Yu and D.L. Kwong, "Oxide-based RRAM: Physical based retention projection", *Proceedings of the European Solid-State Device Research Conference (ESSDERC),* pp.392-395, (2010).

[9] S. Deora, G. Bersuker, M.G. Sung, D.C. Gilmer, P.D. Kirsch, H.F. Li, H. Chong and S. Gausepohl, "Statistical assessment of endurance degradation in high and low resistive states of the HfO_2-based RRAM", *IEEE International Reliability Physics Symposium (IRPS),* pp.MY.2.1-MY.2.5, (2013).

[10] N. Raghavan, K.L. Pey, D.D. Frey and M. Bosman, "Stochastic failure model for endurance degradation in vacancy modulated HfO_x RRAM using the percolation cell framework", *IEEE International Reliability Physics Symposium (IRPS),* pp.MY.9.1-MY.9.7, (2014).

[11] B. Gao, H. Zhang, B. Chen, L. Liu, X. Liu, R. Han, J.F. Kang, Z. Fang, H.Y. Yu, B. Yu and D.L. Kwong, "Modeling of retention failure behavior in bipolar oxide-based resistive switching memory", *IEEE Electron Device Letters,* Vol. 32, No. 3, pp.276-278, (2011).

[12] W.C. Luo, J.C. Liu, Y.C. Lin, C.L. Lo, J.J. Huang, K.L. Lin and T.H. Hou, "Statistical model and rapid prediction of RRAM SET speed–disturb dilemma", *IEEE Transactions on Electron Devices,* Vol. 60, No. 11, pp.3760-3766, (2013).

[13] R. Degraeve, A. Fantini, N. Raghavan, Y.Y. Chen, L. Goux, S. Clima, S. Cosemans, B. Govoreanu, D.J. Wouters, P. Roussel, G.S. Kar, G. Groeseneken and M. Jurczak, "Modeling RRAM set/reset statistics resulting in guidelines for optimized operation", *Symposium on VLSI Technology (VLSIT),* pp.T98-T99, (2013).

[14] R. Degraeve, A. Fantini, S. Clima, B. Govoreanu, L. Goux, Y.Y. Chen, D.J. Wouters, P.J. Roussel, G.S. Kar, G. Pourtois, S. Cosemans, J.A. Kittl, G. Groeseneken, M. Jurczak and L. Altimime, "Dynamic 'hour glass' model for SET and RESET in HfO_2 RRAM", *Symposium on VLSI Technology (VLSIT),* pp.75-76, (2012).

[15] R. Degraeve, L. Goux, S. Clima, B. Govoreanu, Y.Y. Chen, G.S. Kar, P.J. Roussel, G. Pourtois, D.J. Wouters, L. Altimime, M. Jurczak, G. Groeseneken and J.A. Kittl, "Modeling and tuning the filament properties in RRAM metal oxide stacks for optimized stable cycling", *International Symposium on VLSI Technology, Systems, and Applications (VLSI-TSA),* pp.1-2, (2012).

[16] B. Butcher, G. Bersuker, L. Vandelli, A. Padovani, L. Larcher, A. Kalantarian, R. Geer and D.C. Gilmer, "Modeling the effects of different forming conditions on RRAM conductive filament stability", *5th IEEE International Memory Workshop (IMW),* pp.52-55, (2013).

[17] D.B. Strukov and R.S. Williams, "Exponential ionic drift: fast switching and low volatility of thin-film memristors", *Applied Physics A*, Vol. 94, No. 3, pp.515-519, (2009).

[18] S. Yu and H.S.P. Wong, "A phenomenological model for the reset mechanism of metal oxide RRAM", *IEEE Electron Device Letters*, Vol. 31, No. 12, pp.1455-1457, (2010).

[19] A.S. Foster, A.L. Shluger and R.M. Nieminen, "Mechanism of interstitial oxygen diffusion in hafnia", *Physical Review Letters*, Vol. 89, No. 22, 225901, (2002).

[20] N. Raghavan, R. Degraeve, L. Goux, A. Fantini, D.J. Wouters, G. Groeseneken and M. Jurczak, "RTN insight to filamentary instability and disturb immunity in ultra-low power switching HfO_x and AlO_x RRAM", *Symposium on VLSI Technology (VLSIT),* pp.T164-T165, (2013).

[21] S. Ulreich and W. Zwerger, "Where is the potential drop in a quantum point contact?", *Superlattices and Microstructures*, Vol. 23, Issues 3–4, pp. 719-730, (1998).

[22] J.W. McPherson and H.C. Mogul, "Underlying physics of the thermochemical E-model in describing low-field time-dependent dielectric breakdown in SiO_2 thin films", *Journal of Applied Physics*, Vol. 84, No. 3, pp.1513-1523, (1998).

[23] G. Bersuker, Y. Jeon and H.R. Huff, "Degradation of thin oxides during electrical stress", *Microelectronics Reliability*, Vol. 41, Issue 12, pp.1923-1931, (2001).

[24] A. Fantini, L. Goux, R. Degraeve, D.J. Wouters, N. Raghavan, G.S. Kar, A. Belmonte, Y.Y. Chen, B. Govoreanu and M. Jurczak, "Intrinsic switching variability in HfO_2 RRAM", *5th IEEE International Memory Workshop (IMW)*, pp.30-33, (2013).

[25] U. Celano, L. Goux, A. Belmonte, K. Opsomer, A. Franquet, A. Schulze, C. Detavernier, O. Richard, H. Bender, M. Jurczak and W. Vandervorst, "Three-dimensional observation of the conductive filament in nanoscaled resistive memory devices", *Nano Letters*, Vol. 14, No. 5, pp.2401-2406, (2014).

[26] X. Wu, D. Cha, M. Bosman, N. Raghavan, D.B. Migas, V.E. Borisenko, X.X. Zhang, K. Li and K.L. Pey, "Intrinsic nanofilamentation in resistive switching", *Journal of Applied Physics*, Vol. 113, Issue 11, 114503, (2013).

[27] M.A. Villena, M.B. González, F. Jiménez-Molinos, F. Campabadal, J.B. Roldán, J. Suñé, E. Romera and E. Miranda, "Simulation of thermal reset transitions in resistive switching memories including quantum effects", *Journal of Applied Physics*, Vol. 115, Issue 21, 214504, (2014).

[28] S. Lin, L. Zhao, J. Zhang, H. Wu, Y. Wang, H. Qian and Z. Yu, "Electrochemical simulation of filament growth and dissolution in conductive-bridging RAM (CBRAM) with cylindrical coordinates", *IEEE International Electron Devices Meeting (IEDM),* pp.26.3.1-26.3.4, (2012).